UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
09/929,516	08/13/2001	Frank Paetzold	EYEM1340	8236

25548 7590 10/28/2003

MARK M. TAKAHASHI
GRAY CARY WARE & FREIDENRICH, LLP
4365 EXECUTIVE DRIVE, SUITE 1100
SAN DIEGO, CA 92121-2133

EXAMINER	
CAO, HUEDUNG X	
ART UNIT	**PAPER NUMBER**
2671	11

DATE MAILED: 10/28/2003

Please find below and/or attached an Office communication concerning this application or proceeding.

Office Action Summary	Application No.	Applicant(s)	
	09/929,516	PAETZOLD ET AL.	
	Examiner	Art Unit	
	Huedung X Cao	2671	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If the period for reply specified above is less than thirty (30) days, a reply within the statutory minimum of thirty (30) days will be considered timely.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133).
- Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on *16 December 2002* .

2a)☐ This action is **FINAL**. 2b)☒ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) *1-16* is/are pending in the application.

4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) *1-16* is/are rejected.

7)☐ Claim(s) _____ is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☐ The specification is objected to by the Examiner.

10)☐ The drawing(s) filed on _____ is/are: a)☐ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

11)☐ The proposed drawing correction filed on _____ is: a)☐ approved b)☐ disapproved by the Examiner.

If approved, corrected drawings are required in reply to this Office action.

12)☐ The oath or declaration is objected to by the Examiner.

Priority under 35 U.S.C. §§ 119 and 120

13)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a)☐ All b)☐ Some * c)☐ None of:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____ .

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

14)☐ Acknowledgment is made of a claim for domestic priority under 35 U.S.C. § 119(e) (to a provisional application).

a) ☐ The translation of the foreign language provisional application has been received.

15)☐ Acknowledgment is made of a claim for domestic priority under 35 U.S.C. §§ 120 and/or 121.

Attachment(s)

1) ☒ Notice of References Cited (PTO-892)
2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
3) ☒ Information Disclosure Statement(s) (PTO-1449) Paper No(s) *6-9* .
4) ☐ Interview Summary (PTO-413) Paper No(s). _____ .
5) ☐ Notice of Informal Patent Application (PTO-152)
6) ☐ Other: .

DETAILED ACTION

Claim Rejections - 35 USC § 103

1. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

2. This application currently names joint inventors. In considering patentability of the claims under 35 U.S.C. 103(a), the examiner presumes that the subject matter of the various claims was commonly owned at the time any inventions covered therein were made absent any evidence to the contrary. Applicant is advised of the obligation under 37 CFR 1.56 to point out the inventor and invention dates of each claim that was not commonly owned at the time a later invention was made in order for the examiner to consider the applicability of 35 U.S.C. 103(c) and potential 35 U.S.C. 102(e), (f) or (g) prior art under 35 U.S.C. 103(a).

3. Claims 1-16 are rejected under 35 U.S.C. 103(a) as being unpatentable over SHAW et al. (6,320,583) in view of CHEN (5,608,839).

As per claim 1, Shaw teaches the claimed "method for generating facial animation values using a sequence of facial image frames and captured audio data of a speaking actor" (Shaw, column 7, lines 1-15), comprising the steps for: "providing a plurality of visual-facial-animation values based on tracking of facial features in the

sequence of facial image frames of the speaking actor" (Shaw, column 11, lines 20-34); "providing a plurality of audio-facial-animation values based on visemes detected using the captured audio voice data of the speaking actor" (Shaw, column 12, lines 48-65); and "combining the plurality of visual facial animation values and the plurality of audio facial animation values to generate output facial animation values for use in facial animation" (Shaw, column 13, lines 19-35). It is noted that Shaw does not teach the capture of the audio data of a speaking actor and its video facial image is in a "synchronous" manner as claimed. Chen teaches that the "synchronously" captured audio data of a speaking actor and its facial image is well known in the art (Chen, column 4, lines 56-57). It would have been obvious at the time the invention was made, in view of the teaching of Chen, to configure Shaw's method as claimed because Shaw's recorded "visemes" morph (column 3, lines 44-50) representing the speaking word and its corresponding visual facial character would be from a video/audio signal of synchronous audio and video data.

Claim 2 adds into claim 1 "the output facial animation values associated with a mouth for a facial animation are based only on the respective mouth-associated values of the plurality of audio facial animation values" which Shaw teaches in column 13, lines 16-18.

Claim 3 adds into claim 1 "the output facial animation values associated with a mouth for a facial animation are based on a weighted average of the respective mouth-associated values of the plurality of visual facial animation values and the respective mouth-associated values of the plurality of audio facial animation values" which Shaw

teaches in column 13, lines 32-34.

Claim 4 adds into claim 1 "the output facial animation values associated with a mouth for a facial animation are based on Kalman filtering of the respective mouth-associated values of the plurality of visual facial animation values and the respective mouth-associated values of the plurality of audio facial animation values" which would have been obvious because Shaw's combining of the basic facial image and its speaking morph can smooth out the transition through a weight average or filtering process such as Kalman filter. Further, Kalman filter are an extremely well known type of filter and one of ordinary skill in the art would have known to use them for their known benefits in the art (Official Notice, see MPEP 2144.03)

Claim 5 adds into claim 1 "detecting whether speech is occurring in the synchronously captured audio voice data of the speaking actor and, while speech is detected as occurring, generating the output facial animation values associated with a mouth based only on the respective mouth-associated values of the plurality of audio facial animation values and, while speech is not detected as occurring, generating the output facial animation values associated with a mouth based only on the respective mouth-associated values of the plurality of visual facial animation values" which would have been obvious because Shaw's morph can be used to add any particular characteristic or quality such as emotion, facial movement, speech expression, ... to the original face (Shaw, column 3, lines 29-34).

Claim 6 adds into claim 1 "the tracking of facial features in the sequence of facial image frames of the speaking actor is performed using bunch graph matching" which

would have been obvious because Shaw's delta-zones are used to matching the different sets of facial data to form the desired human face (Shaw, column 7, line 54 to column 8, line 6).

Claim 7 adds into claim1 "the tracking of facial features in the sequence of facial image frames of the speaking actor is performed using transformed facial image frames generated based on wavelet transformations" which would have been obvious because Shaw facial animation are based on a basic facial image shape and the additional details of facial features which can be represented as wavelet transformation characteristics because wavelet transforms are an extremely conventional way to transform and present data (Official Notice, see MPEP 2144.03).

Claim 8 adds into claim 1 "tracking of facial features in the sequence of facial image frames of the speaking actor is performed using transformed facial image frames generated based on Gabor wavelet transformations" which would have been obvious because Shaw facial animation are based on a basic facial image shape and the additional details of facial features which can be represented as wavelet transformation characteristics such as Gabor wavelet transformation because wavelet transforms are an extremely conventional way to transform and present data (Official Notice, see MPEP 2144.03).

As per claim 9, Shaw teaches the claimed "apparatus for generating facial animation values using a sequence of facial image frames and captured audio data of a speaking actor" (Shaw, column 7, lines 1-15), comprising the steps for: "means for providing a plurality of visual-facial-animation values based on tracking of facial features

in the sequence of facial image frames of the speaking actor" (Shaw, column 11, lines 20-34); "means for providing a plurality of audio-facial-animation values based on visemes detected using the captured audio voice data of the speaking actor" (Shaw, column 12, lines 48-65); and "means for combining the plurality of visual facial animation values and the plurality of audio facial animation values to generate output facial animation values for use in facial animation" (Shaw, column 13, lines 19-35). It is noted that Shaw does not teach the capture of the audio data of a speaking actor and its video facial image is in a "synchronous" manner as claimed. Chen teaches that the "synchronously" captured audio data of a speaking actor and its facial image is well known in the art (Chen, column 4, lines 56-57). It would have been obvious at the time the invention was made, in view of the teaching of Chen, to configure Shaw's apparatus as claimed because Shaw's recorded "visemes" morph (column 3, lines 44-50) representing the speaking word and its corresponding visual facial character would be from a video/audio signal of synchronous audio and video data.

Claim 10 adds into claim 9 "the output facial animation values associated with a mouth for a facial animation are based only on the respective mouth-associated values of the plurality of audio facial animation values" which Shaw teaches in column 13, lines 16-18.

Claim 11 adds into claim 9 "the output facial animation values associated with a mouth for a facial animation are based on a weighted average of the respective mouth-associated values of the plurality of visual facial animation values and the respective mouth-associated values of the plurality of audio facial animation values" which Shaw

teaches in column 13, lines 32-34.

Claim 12 adds into claim 9 "the output facial animation values associated with a mouth for a facial animation are based on Kalman filtering of the respective mouth-associated values of the plurality of visual facial animation values and the respective mouth-associated values of the plurality of audio facial animation values" which would have been obvious because Shaw's combining of the basic facial image and its speaking morph can smooth out the transition through a weight average or filtering process such as Kalman filter. Further, Kalman filter are an extremely well known type of filter and one of ordinary skill in the art would have known to use them for their known benefits in the art (Official Notice, see MPEP 2144.03).

Claim 13 adds into claim 9 "means for detecting whether speech is occurring in the synchronously captured audio voice data of the speaking actor and, while speech is detected as occurring, generating the output facial animation values associated with a mouth based only on the respective mouth-associated values of the plurality of audio facial animation values and, while speech is not detected as occurring, generating the output facial animation values associated with a mouth based only on the respective mouth-associated values of the plurality of visual facial animation values" which would have been obvious because Shaw's morph can be used to add any particular characteristic or quality such as emotion, facial movement, speech expression, ... to the original face (Shaw, column 3, lines 29-34).

Claim 14 adds into claim 9 "the tracking of facial features in the sequence of facial image frames of the speaking actor is performed using bunch graph matching"

which would have been obvious because Shaw's delta-zones are used to matching the different sets of facial data to form the desired human face (Shaw, column 7, line 54 to column 8, line 6).

Claim 15 adds into claim 9 "the tracking of facial features in the sequence of facial image frames of the speaking actor is performed using transformed facial image frames generated based on wavelet transformations" which would have been obvious because Shaw facial animation are based on a basic facial image shape and the additional details of facial features which can be represented as wavelet transformation characteristics because wavelet transforms are an extremely conventional way to transform and present data (Official Notice, see MPEP 2144.03).

Claim 16 adds into claim 9 "the tracking of facial features in the sequence of facial image frames of the speaking actor is performed using transformed facial image frames generated based on Gabor wavelet transformations" which would have been obvious because Shaw facial animation are based on a basic facial image shape and the additional details of facial features which can be represented as wavelet transformation characteristics such as Gabor wavelet transformation because wavelet transforms are an extremely conventional way to transform and present data (Official Notice, see MPEP 2144.03).

Conclusion

4. The prior art made of record and not relied upon is considered pertinent to applicant's disclosure.

Cosatto et al. (US 6504546 B1) teach A method for modeling three-dimensional objects to create photo-realistic animations using a data-driven approach. The three-dimensional object is defined by a set of separate three-dimensional planes, each plane enclosing an area of the object that undergoes visual changes during animation. Recorded video is used to create bitmap data to populate a database for each three-dimensional plane. The video is analyzed in terms of both rigid movements (changes in pose) and plastic deformation (changes in expression) to create the bitmaps. The modeling is particularly well-suited for animations of a human face, where an audio track generated by a text-to-speech synthesizer can be added to the animation to create a photo-realistic "talking head".

Merrill et al. (US 6181351 B1) teach The animation of a speaking character is synchronized with recorded speech by creating and playing a linguistically enhanced sound file. A sound editing tool employs a speech recognition engine to create the linguistically enhanced sound file from recorded speech and a text of the speech. The speech recognition engine provides timing information related to word breaks and phonemes that is used by the sound editing tool to annotate the speech sound data when creating the linguistically enhanced sound file. When the linguistically enhanced sound file is played to produce sound output, the timing information is retrieved to control the animated character's mouth movement and word pacing in the character's

word balloon. The sound editing tool additionally provides editing functions for manipulating the timing information. A text to speech engine can use the same programming interface as the linguistically enhanced sound file player to send notifications to the animation, providing prototyping without recorded speech. Since both use the same interface, recorded speech can be incorporated at a later time with minimal modifications.

Inquires

Any inquiry concerning this communication or earlier communications from the examiner should be directed to **Huedung Cao** whose telephone number is **(703) 308-5024.**

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, **Mark Zimmerman,** can be reached at **(703) 305-9798**.

Any response to this action should be mailed to:

Commissioner of Patents and Trademarks

Washington, D.C. 20231

or faxed to:

(703) 872-9314 (for Technology Center 2600 only)

Hand-delivered responses should be brought to Crystal Park II, 2121 Crystal Drive, Arlington, VA, Sixth Floor (Receptionist).

Any inquiry of a general nature or relating to the status of this application or proceeding should be directed to the Technology Center 2600 Customer Service Office whose telephone number is (703) 305-0377.

Huedung Cao
Patent Examiner

STEVEN SARAS
SUPERVISORY PATENT EXAMINER
TECHNOLOGY CENTER 2600